Coding and Information Theory
Chapter 3
Entropy (C)

Xuejun Liang

This is the third lecture of chapter 3

Chapter 3: Entropy

3.1 Information and Entropy

3.2 Properties of the Entropy Function

3.3 Entropy and Average Word-length

3.4 Shannon-Fane Coding

3.5 Entropy of Extensions and Products

3.6 Shannon's First Theorem

3.7 An Example of Shannon's First Theorem

Quick Review of Last Lecture

Theorem 3.11: If C is any uniquely decodable r-ary code for a source S, then $L(C) \geq H_r(S)$.

Corollary 3.12: $L(C) = H_r(S)$ if and only if $log_r(p_i)$ is an integer for each i , that is, each $p_i = r^{e_i}$ for some integer $e_i \leq 0$

Efficiency $\eta = \dfrac{H_r(S)}{L(C)}$,

Redundancy $\bar{\eta} = 1 - \eta.$

A Shannon-Fano code C for S has word lengths $l_i = \lceil \log_r(1/p_i) \rceil$

Theorem 3.16: Every r-ary Shannon-Fano code C for a source S satisfies

$$H_r(S) \leq L(C) \leq 1 + H_r(S)$$

- Example 3.18
  - Let S have 5 symbols, with probabilities p_i= 0.3, 0.2, 0.2, 0.2, 0.1 as in Example 2.5
  - Compute Shannon-Fano code word length l_i, $L(C)$, η.
  - Compare with Huffman code.

Compute word length l_i of Shannon-Fano Code

$$l_i = \lceil \log_2(1/p_i) \rceil = \min\{n \in \mathbf{Z} \mid 2^n \geq 1/p_i\}$$

$\lceil lg\ ^1/_{p_i} \rceil = l_i \;\Rightarrow\; lg\ ^1/_{p_i} \leq l_i \;\Rightarrow\; ^1/_{p_i} \leq 2^{l_i}$

- Example 3.18
  - Let S have 5 symbols, with probabilities p_i= 0.3, 0.2, 0.2, 0.2, 0.1 as in Example 2.5
  - Compute Shannon-Fano code word length l_i, $L(C)$, η.
  - Compare with Huffman code.

p_i	3	2	2	2	.1
l_i	2	3	3	3	4

- Example 3.19
  - If p_1 = 1 and p_i = 0 for all i > 1, then $H_r(S) = 0$. An r-ary optimal code D for S has average word-length $L(D) = 1$, so here the upper bound 1 + $H_r(S)$ is attained.

Theorem 3.16: Every r-ary Shannon-Fano code C for a source S satisfies

$$H_r(S) \leq L(C) \leq 1 + H_r(S)$$

3.5 Entropy of Extensions and Products

- Recall from §2.6
  - S^n has q^n symbols $s_{i_1} \dots s_{i_n}$ with probabilities $p_{i_1} \dots p_{i_n}$.
- Theorem 3.20
  - If S is any source then $H_r(S^n) = nH_r(S)$.
- Lemma 3.21
  - If S and T are independent sources then $H_r(S \times T) = H_r(S) + H_r(T)$
- Corollary 3.22
  - If $S_1, \dots, S_n$ are independent sources then
    $H_r(\mathcal{S}_1 \times \cdots \times \mathcal{S}_n) = H_r(\mathcal{S}_1) + \cdots + H_r(\mathcal{S}_n)$

- Lemma 3.21
  - If S and T are independent sources then $H_r(S \times T) = H_r(S) + H_r(T)$

Proof

Independence gives $\Pr(s_i t_j) = p_i q_j$, so

$$\begin{aligned}
H_r(\mathcal{S} \times \mathcal{T}) &= -\sum_i \sum_j p_i q_j \log_r p_i q_j \\
&= -\sum_i \sum_j p_i q_j (\log_r p_i + \log_r q_j) \\
&= -\sum_i \sum_j p_i q_j \log_r p_i - \sum_i \sum_j p_i q_j \log_r q_j \\
&= \Big(-\sum_i p_i \log_r p_i\Big)\Big(\sum_j q_j\Big) + \Big(\sum_i p_i\Big)\Big(-\sum_j q_j \log_r q_j\Big) \\
&= H_r(\mathcal{S}) + H_r(\mathcal{T})
\end{aligned}$$

since $\sum p_i = \sum q_j = 1$.

3.6 Shannon's First Theorem

- Theorem 3.23
  - By encoding S^n with n sufficiently large, one can find uniquely decodable r-ary encodings of a source S with average word-lengths arbitrarily close to the entropy $H_r(S)$.
- Recall that
  - if a code for S^n has average word-length L_n, then as an encoding of S it has average word-length L_n/n.
- Note that
  - the encoding process of S^n for a large n are complicated and time-consuming.
  - the decoding process involves delays

Proof of Shannon's First Theorem

- Theorem 3.23
  - By encoding S^n with n sufficiently large, one can find uniquely decodable r-ary encodings of a source S with average word-lengths arbitrarily close to the entropy $H_r(S)$.

Proof: By Corollary 3.17, $$H_r(S^n) \leq L_n \leq 1 + H_r(S^n)\,,$$

Theorem 3.20 gives $$nH_r(S) \leq L_n \leq 1 + nH_r(S)\,.$$

Dividing by n we get $$H_r(S) \leq \frac{L_n}{n} \leq \frac{1}{n} + H_r(S)\,,$$

So $$\lim_{n\to\infty} \frac{L_n}{n} = H_r(S)\,.$$

3.7 An Example of Shannon's First Theorem

Let S be a source with two symbols s_1, s_2 of probabilities $p_i = 2/3, 1/3$, as in Example 3.2.

- In §3.1, we have $\boxed{H_2(S) = \log_2 3 - \frac{2}{3} \approx 0.918}$
- In §2.6, using binary Huffman codes for S^n with n = 1, 2 and 3, we have $\boxed{L_n/n \approx 1,\ 0.944 \text{ and } 0.938}$
- For larger n it is simpler to use Shannon-Fano codes, rather than Huffman codes.
    - Compute L_n for S^n

$$\boxed{L_n = a_n - \frac{2n}{3}} \qquad \boxed{a_n = \lceil n log_2 3 \rceil}$$

    - Verify $L_n/n \to H_2(S)$

Verify $L_n/n \to H_2(S)$

$$H_2(S) = \log_2 3 - \tfrac{2}{3} \approx 0.918$$

$$L_n = a_n - \frac{2n}{3}$$

$$a_n = \lceil n log_2 3 \rceil$$

$$\frac{L_n}{n} = \frac{a_n}{n} - \frac{2}{3} = \frac{\lceil n \log_2 3 \rceil}{n} - \frac{2}{3}.$$

$$n \log_2 3 \leq \lceil n \log_2 3 \rceil < 1 + n \log_2 3,$$

$$\log_2 3 \leq \frac{\lceil n \log_2 3 \rceil}{n} < \frac{1}{n} + \log_2 3,$$

$$\frac{\lceil n \log_2 3 \rceil}{n} \to \log_2 3$$

$$L_n/n \to H_2(S)$$

Compute L_n for S^n-- (1)

$$L_n = a_n - \frac{2n}{3}$$

$$a_n = \lceil n log_2 3 \rceil$$

S has two symbols s_1, s_2 of probabilities $p_i = 2/3, 1/3$

S^n has 2^n symbols, each consisting of a block of *n* symbols s_1 or s_2

Assume $s \in S^n$ with *k* symbols s_1 and (*n*-*k*) symbols s_2

Then *s* has probability

$$\Pr(\mathbf{s}) = \left(\frac{2}{3}\right)^k \left(\frac{1}{3}\right)^{n-k} = \frac{2^k}{3^n}.$$

The symbol *s* has a **Shannon-Fano** code-word of length

$$l_k = \left\lceil \log_2\left(\frac{1}{\Pr(\mathbf{s})}\right)\right\rceil = \left\lceil \log_2\left(\frac{3^n}{2^k}\right)\right\rceil = \lceil n\log_2 3 - k\rceil = a_n - k,$$

Compute L_n for S^n -- (2)

$$L_n = a_n - \frac{2n}{3}$$

$$a_n = \lceil n\log_2 3 \rceil$$

For each k = 0, 1, ..., n, the number of such symbols s is $C(k, n)$

Hence the average word-length (for encoding S^n) is

$$\begin{aligned} L_n &= \sum_{k=0}^{n} \binom{n}{k} \Pr(\mathbf{s})\, l_k \\ &= \sum_{k=0}^{n} \binom{n}{k} \frac{2^k}{3^n}(a_n - k) \\ &= \frac{1}{3^n}\left(a_n \sum_{k=0}^{n} \binom{n}{k} 2^k - \sum_{k=0}^{n} k \binom{n}{k} 2^k\right) \end{aligned}$$

(3.9)

By the Binomial Theorem

$$(1+x)^n = \sum_{k=0}^{n} \binom{n}{k} x^k \qquad (3.10)$$

$\big\downarrow$ X = 2

$$\sum_{k=0}^{n} \binom{n}{k} 2^k = 3^n.$$

Compute L_n for S^n -- (3)

$$L_n = a_n - \frac{2n}{3}$$

$$a_n = \lceil n \log_2 3 \rceil$$

Differentiating (3.10) and then multiplying by *x*, we have

$$nx(1+x)^{n-1} = \sum_{k=1}^{n} k \binom{n}{k} x^k = \sum_{k=0}^{n} k \binom{n}{k} x^k,$$

X = 2

$$\sum_{k=0}^{n} k \binom{n}{k} 2^k = 2n.3^{n-1}.$$

By the Binomial Theorem

$$(1+x)^n = \sum_{k=0}^{n} \binom{n}{k} x^k \qquad (3.10)$$

X = 2

$$\sum_{k=0}^{n} \binom{n}{k} 2^k = 3^n.$$

Substituting in (3.9), we have

$$L_n = \frac{1}{3^n}\left(a_n 3^n - 2n.3^{n-1}\right) = a_n - \frac{2n}{3}$$